

The private capacity of quantum channels is not additive

Ke Li,^{1,*} Andreas Winter,^{2,3,†} XuBo Zou,^{1,‡} and GuangCan Guo^{1,§}

¹*Key Laboratory of Quantum Information, University of Science and Technology of China, Chinese Academy of Sciences, Hefei, Anhui 230026, China*

²*Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, U.K.*

³*Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, Singapore 117542*

(Dated: 30 August 2009)

Recently there has been considerable activity on the subject of additivity of various quantum channel capacities. Here, we construct a family of channels with sharply bounded classical, hence private capacity. On the other hand, their quantum capacity when combined with a zero private (and zero quantum) capacity erasure channel, becomes larger than the previous classical capacity. As a consequence, we can conclude for the first time that the classical private capacity is non-additive. In fact, in our construction even the quantum capacity of the tensor product of two channels can be greater than the sum of their individual classical private capacities. We show that this violation occurs quite generically: every channel can be embedded into our construction, and a violation occurs whenever the given channel has larger entanglement assisted quantum capacity than (unassisted) classical capacity.

Information Theory, established by Claude Shannon in the 1940s as a “Mathematical Theory of Communication” [1], is the theoretical foundation of today’s communication technologies. The main problem it is concerned with is how much information can be transmitted down a noisy channel asymptotically, i.e. the capacity of the channel. Shannon provided a beautifully simple formula for the capacity of a discrete memoryless channel, which only involves an entropic expression of a single channel use. Subsequent research revealed that this simple capacity formula fully characterizes the information-carrying capability of a channel under a large range of circumstances [2], serving as a very robust measure. E.g., the ability of two channels together to transmit information is quantified by the sum of their individual capacities.

Our world however is not the classical one of Shannon’s noisy channels, but is at a basic level described by quantum theory. To understand the ultimate limit the laws of physics impose on our ability to communicate, the underlying quantum behavior of the channels should be considered. A quantum channel \mathcal{N} is mathematically described by an isometric map V from the input Hilbert space A to the combined Hilbert space of the output B and the so-called environment system E . Then the channel and its natural complement $\tilde{\mathcal{N}}$ act as

$$\mathcal{N}(\rho) = \text{Tr}_E V \rho V^\dagger, \quad \tilde{\mathcal{N}}(\rho) = \text{Tr}_B V \rho V^\dagger.$$

It can in general not only convey classical messages, but also quantum data, i.e. a Hilbert space of quantum states. It can also carry classical private information, inaccessible to the environment, enabling the classically impossible, provably unconditionally secure key distribution [3]. Naturally, deriving capacity formulae of a quantum channel for transmitting various kinds of information is a central task of quantum information theory.

The classical capacity, $C(\mathcal{N})$, is the maximal rate of classical information that the quantum channel \mathcal{N} can

asymptotically transmit with vanishing errors. In contrast to the classical capacity, the definition of classical private capacity $P(\mathcal{N})$ further requires that the classical information conveyed is secret from the environment. Finally, the quantum capacity $Q(\mathcal{N})$ quantifies how large a Hilbert space of states the channel \mathcal{N} can transmit asymptotically and with the error approaching zero. Operationally, quantum information transmission implies private classical transmission, which in turn implies plain classical communication. I.e.,

$$C(\mathcal{N}) \geq P(\mathcal{N}) \geq Q(\mathcal{N}). \quad (1)$$

Despite considerable progress, tractable formulae for the quantum, private and unrestricted classical capacities are still out of reach. The HSW theorem [4], Devetak [5] and the LSD theorem [5, 6, 7] give the classical, private and quantum capacities, respectively, as the regularisation of single-letter quantities:

$$\chi(\mathcal{N}) \leq C(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}), \quad (2)$$

$$P^{(1)}(\mathcal{N}) \leq P(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} P^{(1)}(\mathcal{N}^{\otimes n}), \quad (3)$$

$$Q^{(1)}(\mathcal{N}) \leq Q(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} Q^{(1)}(\mathcal{N}^{\otimes n}). \quad (4)$$

All three single-letter quantities are obtained via finite optimizations of entropic expressions: the *Holevo capacity* $\chi(\mathcal{N})$ is the maximum over all ensembles $\{p_i, \rho_i\}$ of states on A of the *Holevo information*

$$\chi_{\{p_i, \rho_i\}}(\mathcal{N}) = H\left(\mathcal{N}\left(\sum_i p_i \rho_i\right)\right) - \sum_i p_i H(\mathcal{N}(\rho_i)), \quad (5)$$

where $H(\rho) = -\text{Tr} \rho \log \rho$ is the von Neumann entropy (\log is always the binary logarithm). Similarly, $P^{(1)}(\mathcal{N}) = \max_{\{p_i, \rho_i\}} (\chi_{\{p_i, \rho_i\}}(\mathcal{N}) - \chi_{\{p_i, \rho_i\}}(\tilde{\mathcal{N}}))$, and

$Q^{(1)}(\mathcal{N}) = \max_{\rho} I_c(\rho, \mathcal{N})$, with the *coherent information* [8]

$$I_c(\rho, \mathcal{N}) = H(\mathcal{N}(\rho)) - H(\tilde{\mathcal{N}}(\rho)). \quad (6)$$

Neither $\chi(\mathcal{N})$, nor $Q^{(1)}(\mathcal{N})$, nor $P^{(1)}(\mathcal{N})$ are additive; in fact, $\chi \neq C$ [9], $P^{(1)} \neq P$ [10], $Q^{(1)} \neq Q$ [11]. However, for certain classes of channels it is known that $C(\mathcal{N}) = \chi(\mathcal{N})$ [12, 13, 14], and for other classes $P(\mathcal{N}) = P^{(1)}(\mathcal{N})$, $Q(\mathcal{N}) = Q^{(1)}(\mathcal{N})$ [15].

As measures of a channel's information transmitting capability, the above three capacity quantities might be expected to be robust, i.e. just like Shannon's capacity for classical channels, to be applicable under a large range of settings. While this is no longer true when various auxiliary resources (e.g. free entanglement or classical communications) are available [16], another weird feature of the quantum capacity was discovered recently. Smith and Yard [17] show that, as a function of channels, $Q(\mathcal{N})$ is not additive. Specifically, for the two channels \mathcal{N}_1 and \mathcal{N}_2 with \mathcal{N}_1 satisfying $Q(\mathcal{N}_1) = 0$ and $P(\mathcal{N}_1) > 0$, and \mathcal{N}_2 the (zero quantum and zero private capacity) 50% erasure channel, $Q(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq \frac{1}{2}P(\mathcal{N}_1) > 0$. One might attribute this superactivation of quantum capacity to the ability to transmit privacy [18], recalling the close relationship between $Q(\mathcal{N})$ and $P(\mathcal{N})$ [19]. But surprisingly again, Smith and Smolin [20] have found two channels such that either they have large joint quantum capacity but negligible individual private classical capacities, or one of them exhibits a large non-additivity of χ .

In this Letter, we present quantum channels $\mathcal{T}_{\mathcal{N}}^k$ for given channel \mathcal{N} with finite environment dimension (this includes all channels with finite dimensional input and output), and integer k ; it inherits input and output from \mathcal{N} , but has also auxiliary registers. We can show that $C(\mathcal{N}) \leq C(\mathcal{T}_{\mathcal{N}}^k) \leq C(\mathcal{N}) + \delta(k)$, where $\delta(k)$ goes to zero as $k \rightarrow \infty$. Regarding the capability of the channel $\mathcal{T}_{\mathcal{N}}^k$, together with a 50% erasure channel \mathcal{A} , for quantum communication, we find that the quantum capacity of the combined channel $\mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A}$ is lower bounded by $Q_E(\mathcal{N})$, the entanglement-assisted quantum capacity of \mathcal{N} [21]. So, for channels \mathcal{N} such that $Q_E(\mathcal{N}) > C(\mathcal{N})$, $\mathcal{T}_{\mathcal{N}}^k$ – when combined with the above erasure channel – can transmit more quantum information than its classical capacity $C(\mathcal{T}_{\mathcal{N}}^k)$. Referring to Eq. (1), we conclusively prove that the classical private capacity, in fact even the quantum capacity, of two channels can be greater than the sum of their individual classical private capacities. Our findings not only demonstrate that the classical private capacity of a quantum channel is generally not additive, but also yield another counterexample to the additivity of quantum capacity, of which the underlying reasoning is different from that of Smith and Yard's [17].

The channel construction. In the Stinespring representation $\mathcal{N}(\rho) = \text{Tr}_E V \rho V^\dagger$, the partial trace embodies all the noise of the channel as loss of information; if Bob

got E as well, there would be no noise at all as he can undo the isometry. However, a well-known way of giving him E anyway, is to completely randomize it: denoting the discrete Weyl operators on E by W_j ($j = 1, \dots, |E|^2$), if the channel internally picks j uniformly at random and applies W_j to E , it creates a new channel with output $\mathcal{N}(\rho)^B \otimes \frac{1}{|E|} \mathbb{1}^E$. The extra register is always constant, so the new channel has the exact same information properties as \mathcal{N} . The idea of the following channel construction is to add another “gadget” on top of this, which outputs some randomness approximating the uniform distribution above – see Fig. 1; so, intuitively, on its own it does not alter too much the classical capacity of the channel, but if paired with the right resources can increase the quantum capacity.

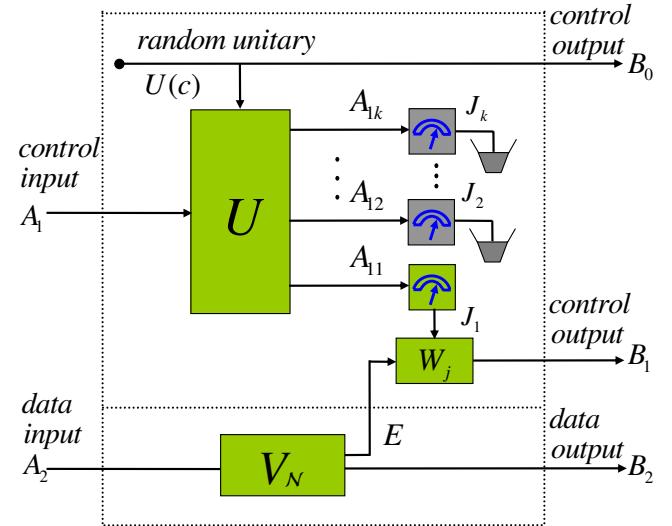


FIG. 1: The channel $\mathcal{T}_{\mathcal{N}}^k$: in the lower part it contains \mathcal{N} (with input register A_2 , output register B_2 , and environment E). It also has another input register A_1 of dimension $c = |E|^{2k}$, which we view in a fixed way as a tensor product of k $|E|^{2k}$ -dimensional systems A_{11}, \dots, A_{1k} , each coming with a fixed computational basis $\{|j\rangle\}_{j=1, \dots, |E|^{2k}}$. This big register is subjected to a random unitary rotation U , where U is chosen from the Haar measure and subsequently output (a classical description of it) in register B_0 . All registers A_{12}, \dots, A_{1k} are discarded, only A_{11} is measured in the computational basis, and the result j used to control a unitary transformation (Weyl operator) W_j on the environment E , which is then output in the register B_1 . A formal definition can be found in the Auxiliary Material [22].

A comment on why we need the rather large register A_1 , most of which is discarded anyway. In fact, the size (parametrized by k) has a double purpose: on the one hand, we need A_{11} to be close to maximally mixed for most inputs. But more importantly, to make it very “costly”, though not impossible, to use entanglement with another system to access the index J_1 (see the proof of Theorem 1).

The additivity violation. Now, if we knew that the Holevo quantity $\chi(\mathcal{T}_{\mathcal{N}}^k)$ were additive for $\mathcal{T}_{\mathcal{N}}^k$, we would have $C(\mathcal{T}_{\mathcal{N}}^k) = \chi(\mathcal{T}_{\mathcal{N}}^k)$. Since it is possible to show that $\chi(\mathcal{T}_{\mathcal{N}}^k) \leq \chi(\mathcal{N}) + o(1)$ – this is a special case of Theorem 1 below –, we would have an upper bound

$$P(\mathcal{T}_{\mathcal{N}}^k) \leq C(\mathcal{T}_{\mathcal{N}}^k) \leq C(\mathcal{N}) + o(1). \quad (7)$$

While we are not able to show additivity for the channels $\mathcal{T}_{\mathcal{N}}^k$, the above relation is nevertheless true. In fact, we have the following general theorem, proved in full in the Auxiliary Material section [22].

Theorem 1 *For any channel \mathcal{N} with input space A , output space B and environment E , and any integer k , let $\delta(k) = \frac{1}{k}(5 + 4 \log |E|)$. Then, for arbitrary channel \mathcal{E} ,*

$$\chi(\mathcal{N} \otimes \mathcal{E}) \leq \chi(\mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{E}) \leq \chi(\mathcal{N} \otimes \mathcal{E}) + \delta(k). \quad (8)$$

As a consequence,

$$C(\mathcal{N}) \leq C(\mathcal{T}_{\mathcal{N}}^k) \leq C(\mathcal{N}) + \delta(k).$$

On the other hand, we can get a lower bound on $P(\mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A})$, where \mathcal{A} is the 50% erasure channel of input dimension c ; note that by the no-cloning principle, $P(\mathcal{A}) = Q(\mathcal{A}) = 0$. Since the private classical capacity is not smaller than the quantum capacity, which is in turn lower bounded by the coherent information, we evaluate the coherent information of the channel $\mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A}$. Let us look at an input state as follows: Alice prepares a maximally entangled state $\Phi^{A_1 C}$ of dimension $c \times c$ and feeds the two halves into the control input (A_1) and the 50% erasure channel (C ; its quantum output we also denote C , and the erasure flag D). She feeds another arbitrary state ρ^{A_2} , whose purification is denoted as $|\varphi\rangle^{AA_2}$, into the data input and keeps the system A . I.e., we compute the coherent information with respect to the input state $\sigma^{A_1 A_2 C} = \Phi^{A_1 C} \otimes \rho^{A_2}$, so that the final state after the channel action is

$$\omega^{AB_0 B_1 B_2 C D} = (\text{id}_A \otimes \mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A}_C)(\Phi^{A_1 C} \otimes \varphi^{AA_2}).$$

The coherent information, with respect to this state, is

$$I_c(\sigma^{A_1 A_2 C}, \mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A}) = H(B_1 B_2 C D | B_0) - H(A B_1 B_2 C D | B_0).$$

By an argument similar to that in [20], we divide the computation into two cases: the information sent into \mathcal{A} is erased or not erased,

$$I_c(\sigma^{A_1 A_2 C}, \mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A}) = \frac{1}{2}(I_c^{\text{erased}} + I_c^{\text{not-erased}}).$$

In the erased case, the state is decoupled between AB_2 and $B_0 B_1 C$, so the coherent information simplifies to

$$\begin{aligned} I_c^{\text{erased}} &= H(B_2) - H(A B_2) \\ &= H(\mathcal{N}(\rho)) - H((\text{id} \otimes \mathcal{N})|\varphi\rangle\langle\varphi|^{AA_2}). \end{aligned}$$

When the transmitted information is not erased, Bob will be able to correct the errors encountered by the noisy channel \mathcal{N} as follows. Bob reads the output B_0 , learning which unitary transformation is applied by the channel $\mathcal{T}_{\mathcal{N}}^k$. Then he can measure C in the proper basis to get j , and then apply W_j^\dagger to B_1 , recovering the environment E . As a result, Bob possesses the output and the environment of \mathcal{N} simultaneously, effectively obtaining the quantum information input into \mathcal{N} completely. In this case, the system $B_0 C$ is decoupled from $A B_1 B_2$, which is in the pure state $(\mathbb{1} \otimes V)|\varphi\rangle^{AA_2}$. So,

$$I_c^{\text{not-erased}} = H(B_1 B_2) - H(A B_1 B_2) = H(\rho^{A_2}).$$

Adding these two cases together, we have

$$I_c(\sigma, \mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A}) = \frac{1}{2} \left[H(\rho) + H(\mathcal{N}(\rho)) - H((\text{id} \otimes \mathcal{N})\varphi) \right].$$

The term in brackets on the right hand side is called quantum mutual information (between input and output of \mathcal{N}). In [21], it is proved that the maximum over ρ of the right hand side is the entanglement-assisted quantum capacity $Q_E(\mathcal{N})$ of the channel \mathcal{N} . I.e.,

$$Q_E(\mathcal{N}) = \max_{\rho} I_c(\Phi \otimes \rho, \mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A}),$$

and hence

$$P(\mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A}) \geq Q(\mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A}) \geq Q_E(\mathcal{N}). \quad (9)$$

Now, comparing Eqs. (7) and (9), also making use of Eq. (1), we see that for all channels \mathcal{N} such that

$$Q_E(\mathcal{N}) > C(\mathcal{N}), \quad (10)$$

we have, for sufficiently large k ,

$$P(\mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A}) \geq Q(\mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{A}) > P(\mathcal{T}_{\mathcal{N}}^k) \geq Q(\mathcal{T}_{\mathcal{N}}^k). \quad (11)$$

Note that the channel \mathcal{A} has zero private classical capacity and zero quantum capacity, so Eq. (11) exhibits the violations of the additivity of private classical capacity and the quantum capacity at the same time.

All we need now is to find quantum channels that satisfy Eq. (10). One example is the depolarizing channel of arbitrary dimension d , for which both capacities are known [13, 21], $\mathcal{D}_q(\rho) = (1 - q)\rho + q\frac{1}{d}\mathbb{1}$. For large d , the gap becomes asymptotically $\frac{1}{2}H(q, 1 - q)$ [22].

There also exist large additivity violations: In [23, Theorem V.1] it is proven that in sufficiently large dimension d , there exist $n = \lfloor (\log d)^4 \rfloor$ orthogonal bases $\mathcal{B}_\nu = (|b_1^{(\nu)}\rangle, \dots, |b_d^{(\nu)}\rangle)$ such that for all states ρ ,

$$\frac{1}{n} \sum_{\nu=1}^n H(\mathcal{B}_\nu | \rho) \geq \log d - 4,$$

where $H(\mathcal{B}_\nu | \rho) = -\sum_{i=1}^d \langle b_i^{(\nu)} | \rho | b_i^{(\nu)} \rangle \log \langle b_i^{(\nu)} | \rho | b_i^{(\nu)} \rangle$ is the Shannon entropy of the outcome distribution when

measuring the state ρ in basis \mathcal{B}_ν . What this means is that the channel \mathcal{N} from d to dn dimensions, defined as

$$\mathcal{N}(\rho) = \sum_{\nu=1}^n \sum_{i=1}^d \frac{1}{n} \langle b_i^{(\nu)} | \rho | b_i^{(\nu)} \rangle |i\rangle\langle i|^B \otimes |\nu\rangle\langle \nu|^{B'},$$

satisfies $\chi(\mathcal{N}) \leq 4$. Since the channel is entanglement-breaking, the additivity result of [14] applies, so $C(\mathcal{N}) = \chi(\mathcal{N}) \leq 4$. On the other hand, it is straightforward to see that $Q_E(\mathcal{N}) = \frac{1}{2} \log d$. Thus we find that almost the entire bandwidth of \mathcal{N} can be activated by the presence of entanglement. Now, to construct the example for activation of the secret capacity by a 50% erasure channel, we observe that $|E| = dn = d(\log d)^4$. We choose $k = \log |E|$, and get from Theorem 1 that $C(\mathcal{T}_\mathcal{N}^k) \leq O(1)$, while at the same time $Q(\mathcal{T}_\mathcal{N}^k \otimes \mathcal{A}) \geq \frac{1}{2} \log d$. Note however that the total input dimension is $2^{O((\log d)^2)}$, which is also the input dimension of the 50% erasure channel.

Conclusion. We showed a way of converting any gap between classical capacity and entanglement-assisted quantum capacity of a channel into a violation of the additivity of the private capacity of the channel tensored with a 50% erasure channel. In fact, the *quantum* capacity of the tensor product channel is larger than the *classical* capacity of the single channel.

The construction is based on a certain embedding of the given channel into a version of the echo-correctable channels from [16]. That the pairing with the erasure channel gives larger quantum capacity follows from the echo-correctable reasoning of the benefit of sharing entanglement. On the other hand, the upper bound on the classical capacity relies on showing that the additional “gadgets” built around the given channel increase the capacity by an arbitrarily small amount. The argument to do so is different from the one proving additivity of χ of the channel (which we cannot do for $\mathcal{T}_\mathcal{N}^k$), and also from the use of the recent continuity bound [24] (which cannot be applied as $\mathcal{T}_\mathcal{N}^k$ is at finite distance from any channel for which we know the capacity).

Thus, we even get a new type of example for the non-additivity of the quantum capacity Q , which is different from Smith and Yard’s [17] as our channel is not PPT entanglement binding. Furthermore, while in [17] the lower bound of half the private capacity on the quantum capacity of the tensor product was enough, here we experience even a large gap between these two quantities. However, we also note a conceptual analogy in the constructions: The PPT entanglement binding channel used in [17] derives from a so-called *pbit state* [25]. It provides Alice and Bob with shared randomness – which is made private by distributing the purification among Alice and Bob, but in a scrambled way that makes it impossible for them to recover much of the entanglement. Our channel randomizes the environment and hence gives it to Bob in an encrypted way, limiting the receiver’s knowledge

about the noise encountered by the channel. In the construction of [17] as in the present one, the availability of additional resources allows Alice and Bob to break the encryption and access the entanglement.

Acknowledgments. KL thanks the University of Bristol and CQT at NUS, where part of this work was done, for their hospitality. KL, XBZ and GCG were supported by the National Fundamental Research Program (Grant No. 2009CB929601), the National Natural Science Foundation of China (Grant Nos. 10674128 and 60121503), and the Innovation Funds and “Hundreds of Talents” program of the Chinese Academy of Sciences and Doctor Foundation of Education Ministry of China (Grant No. 20060358043). AW was supported by the U.K. EPSRC, by the Royal Society, a Philip Leverhulme Prize, and by the European Commission. The Centre for Quantum Technologies is funded by the Singapore Ministry of Education and the National Research Foundation as part of the Research Centres of Excellence programme.

* Electronic address: leeke@mail.ustc.edu.cn

† Electronic address: a.j.winter@bris.ac.uk

‡ Electronic address: xbz@ustc.edu.cn

§ Electronic address: gcg@ustc.edu.cn

- [1] C. E. Shannon, Bell Syst. Tech. J. **27**:379 (1948).
- [2] T. M. Cover, J. A. Thomas, *Elements of Information Theory* (New York: Wiley, 1991).
- [3] C. H. Bennett, G. Brassard, Proc. IEEE Int. Conf. Computers, Syst. and Signal Proc., pp. 175 (1984).
- [4] A. S. Holevo, IEEE Trans. Inf. Theory **44**:269 (1998); B. Schumacher, M. D. Westmoreland, Phys. Rev. A **56**:131 (1997).
- [5] I. Devetak, IEEE Trans. Inf. Theory **51**:44 (2005).
- [6] S. Lloyd, Phys. Rev. A **55**:1613 (1997).
- [7] P. W. Shor, <http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/>.
- [8] B. Schumacher, M. Nielsen, Phys. Rev. A **54**:2629 (1996).
- [9] M. Hastings, Nature Physics **5**:255 (2009).
- [10] G. Smith, J. Renes, J. A. Smolin, Phys. Rev. Lett. **100**:170502 (2008).
- [11] P. W. Shor, J. A. Smolin, arXiv:quant-ph/9604006 (1996); D. P. DiVincenzo, P. W. Shor, J. A. Smolin, Phys. Rev. A **57**(2):830 (1998).
- [12] C. King, J. Math. Phys. **43**:4641 (2002).
- [13] C. King, IEEE Trans. Inf. Theory **49**(1):221 (2003).
- [14] P. W. Shor, J. Math. Phys. **43**:4334 (2002).
- [15] I. Devetak, P. W. Shor, Comm. Math. Phys. **256**(2):287 (2005).
- [16] C. H. Bennett, I. Devetak, P. W. Shor, J. A. Smolin, Phys. Rev. Lett. **96**:150502 (2006).
- [17] G. Smith, J. Yard, Science **321**:1812 (2008).
- [18] J. Oppenheim, Science **321**:1783 (2008).
- [19] I. Devetak, A. Winter, Phys. Rev. Lett. **93**:080501 (2004).
- [20] G. Smith, J. A. Smolin, Phys. Rev. Lett. **102**:010501 (2009).
- [21] C. H. Bennett, P. W. Shor, J. A. Smolin, A. V. Thapliyal,

Phys. Rev. Lett. **83**:3081 (1999) and IEEE Trans. Inf. Theory **48**:2637 (2002).

[22] See EPAPS documents for mathematical details.

[23] P. Hayden, D. Leung, P. W. Shor, A. Winter, Comm. Math. Phys. **250**:371 (2004).

[24] D. Leung, G. Smith, arXiv[quant-ph]:0810.4931 (2008).

[25] K. Horodecki, M. Horodecki, P. Horodecki, J. Oppenheim, Phys. Rev. Lett. **94**:160502 (2005).

[26] Ch. Dankert, R. Cleve, J. Emerson, E. Livine, arXiv:quant-ph/0606161 (2006).

[27] D. Gross, K. Audenaert, J. Eisert, J. Math. Phys. **48**:052104 (2007).

[28] G. Lindblad, Comm. Math. Phys. **40**:147 (1975); A. Uhlmann, Comm. Math. Phys. **54**:21 (1977).

THE CHANNEL CONSTRUCTION (AUXILIARY MATERIAL)

Mathematically, the channel depicted in Fig. 1 is written

$$\begin{aligned} T_N^k(\rho^{A_1 A_2}) = & \int_U dU [U]^{B_0} \otimes \sum_j (W_j^{E \rightarrow B_1})(V^{A_2 \rightarrow B_2 E}) \\ & (\text{Tr}_{A_1} [U^{A_1} \rho U^{A_1 \dagger} |j\rangle\langle j|^{A_1}]) (V^{A_2 \rightarrow B_2 E})^\dagger (W_j^{E \rightarrow B_1})^\dagger, \end{aligned} \quad (12)$$

where the notation $[U]$ denotes a classical label realized as mutually orthogonal states. If there are only countably many values of U , these may be thought of as orthogonal projectors $|U\rangle\langle U|$ on an appropriate Hilbert space.

PROOF OF THEOREM 1 (AUXILIARY MATERIAL)

The left inequality in Eq. (8) is trivial (simply ignore the registers A_1 , B_0 and B_1); also, once the upper bound is proved, the inequality for the capacity follows by induction. Hence we concentrate on the right inequality in (8). We shall need a number of auxiliary results.

Lemma 2 Consider an arbitrary state ρ on $\mathbb{C}^r \otimes \mathbb{C}^s$, and a fixed basis $\{|j\rangle\}_{j=1,\dots,r}$ of \mathbb{C}^r . Then for a Haar-distributed random unitary $U \in \mathcal{U}(\mathbb{C}^r \otimes \mathbb{C}^s)$, the random variable J is defined as follows:

$$\Pr\{J = j|U\} = \langle j| \text{Tr}_s U \rho U^\dagger |j\rangle,$$

and it holds that

$$\mathbb{E}_U H(J|U) \geq \log r - \log \left(1 + \frac{1}{s}\right) \geq \log r - \frac{2}{s}.$$

Proof. Without loss of generality, the input state is some fixed pure state, so that after the unitary and before the measurement, we have a uniformly distributed random state $|\phi\rangle$. Using the bound $H(J) = -\sum_j P_J(j) \log P_J(j) \geq -\log \sum_j P_J(j)^2$ and the concavity of $\log x$, we get now by symmetry

$$\begin{aligned} \mathbb{E}_U H(J|U) & \geq -\log \mathbb{E}_\phi \sum_{j=1}^r \langle j| \text{Tr}_s \phi |j\rangle^2 \\ & = -\log (r \mathbb{E}_\phi \langle 1| \text{Tr}_s \phi |1\rangle^2) \\ & = -\log \left(r \mathbb{E}_\phi [\text{Tr} \phi (|1\rangle\langle 1|^r \otimes \mathbb{1}^s)]^2\right). \end{aligned} \quad (13)$$

For the latter expectation, we use a well-known trick:

$$\begin{aligned} & \mathbb{E}_\phi [\text{Tr} \phi (|1\rangle\langle 1|^r \otimes \mathbb{1}^s)]^2 \\ & = \mathbb{E}_\phi \text{Tr} ((\phi \otimes \phi) (|1\rangle\langle 1|^r \otimes \mathbb{1}^s \otimes |1\rangle\langle 1|^r \otimes \mathbb{1}^s)) \\ & = \text{Tr} \left(\frac{\mathbb{1} + F}{rs(rs+1)} (|1\rangle\langle 1|^r \otimes \mathbb{1}^s \otimes |1\rangle\langle 1|^r \otimes \mathbb{1}^s) \right), \end{aligned}$$

where we have introduced a second tensor copy of the total Hilbert space $\mathbb{C}^r \otimes \mathbb{C}^s$, and F is the SWAP operator of the two. The last line evaluates easily to

$$\begin{aligned}\mathbb{E}_\phi [\text{Tr } \phi(|1\rangle\langle 1|^r \otimes \mathbb{1}^s)]^2 &= \frac{1}{rs(rs+1)}(s^2 + s) \\ &= \frac{1}{r^2} \frac{s+1}{s+1/r} \leq \frac{1}{r^2} \frac{s+1}{s}.\end{aligned}$$

Inserting this into Eq. (13) concludes the proof. \square

Remark 3 From the above calculation we see that the full unitary invariance of the Haar measure is not required; we only need to be able to perform the average purity of the reduced state, which is a quadratic function of the random pure state φ . Thus, it is sufficient to draw U from a so-called unitary 2-design [26]; in prime power dimension it is known that the Clifford group is an example of a finite 2-design, see e.g. [27].

Now, in the channel we imagine that, just as A_{11} , also the $k-1$ registers A_{12}, \dots, A_{1k} are measured in a fixed basis, resulting in k random variables J_1, \dots, J_k . For simplicity, sometimes we write $J_\ell \dots J_k$ as J_ℓ^k , and similarly for the measurement results, j_ℓ . The variables J_2, \dots, J_k are being traced over, so the channel does not change (See Fig. 1).

Lemma 4 For any state $\sigma^{A_1 A_2 B_3}$, suppose we feed A_1 into the channel \mathcal{T}_N^k but keep A_2 unchanged, let $\omega^{B_0 J_1^k A_2 B_3}$ be the state after applying the random unitary U and then doing measurements on $A_{1\ell}$, i.e.

$$\begin{aligned}\omega^{B_0 J_1^k A_2 B_3} &= \int_U dU [U]^{B_0} \otimes \sum_{j_1 \dots j_k} \bigotimes_{\ell=1}^k [j_\ell]^{J_\ell} \\ &\otimes \langle j_1 \dots j_k | U^{A_1} \sigma^{A_1 A_2 B_3} U^{A_1 \dagger} | j_1 \dots j_k \rangle.\end{aligned}$$

Then we have

$$\begin{aligned}I(J_1; A_2 B_3 | B_0) &\leq \frac{1}{k} I(J_1^k; A_2 B_3 | B_0) \\ &+ \frac{1}{k} \sum_{\ell=1}^{k-1} I(J_\ell; J_{\ell+1}^k | B_0),\end{aligned}\tag{14}$$

where

$$\begin{aligned}I(X : Y) &= H(X) + H(Y) - H(XY) \\ &= H(\rho_X) + H(\rho_Y) - H(\rho_{XY})\end{aligned}$$

is the (quantum) mutual information between two subsystems of a bipartite state ρ_{XY} with marginals ρ_X and ρ_Y , and the informations conditional on B_0 are averages over the classical states of this register.

Proof. We use the chain rule to get

$$\begin{aligned}I(J_1^k; A_2 B_3 | B_0) &= I(J_1; A_2 B_3 | B_0) + I(J_2^k; A_2 B_3 | B_0 J_1) \\ &= I(J_1; A_2 B_3 | B_0) + I(J_2^k; A_2 B_3 J_1 | B_0) \\ &\quad - I(J_1; J_2^k | B_0) \\ &\geq I(J_2^k; A_2 B_3 | B_0) \\ &\quad + I(J_1; A_2 B_3 | B_0) - I(J_1; J_2^k | B_0).\end{aligned}$$

Iterating this step for $I(J_2^k; A_2 B_3 | B_0)$, then $I(J_3^k; A_2 B_3 | B_0)$, etc., we obtain

$$\begin{aligned}I(J_1^k; A_2 B_3 | B_0) &\geq \sum_{\ell=1}^k I(J_\ell; A_2 B_3 | B_0) \\ &\quad - \sum_{\ell=1}^{k-1} I(J_\ell; J_{\ell+1}^k | B_0).\end{aligned}\tag{15}$$

By symmetry all the $I(J_\ell; A_2 B_3 | B_0)$ are the same and equal to $I(J_1; A_2 B_3 | B_0)$, concluding the proof. \square

Now, to show Eq. (8), we need to do the following. Given another channel \mathcal{E} , whose input and output registers we denote A_3 and B_3 , respectively, we first have to show that correlated inputs between A_1 and $A_2 A_3$ are (almost) of no use, and hence that the control part of our channel \mathcal{T}_N^k can be skipped.

Mathematically, we have to look at a given ensemble of input states $\{\lambda_x, \varphi_x^{A_1 A_2 A_3}\}$ for $\mathcal{T}_N^k \otimes \mathcal{E}$. We shall find a new ensemble of states only on $A_2 A_3$ which has almost the same Holevo information, even when we consider only the output registers $B_2 B_3$, i.e. those of $\mathcal{N} \otimes \mathcal{E}$. It turns out that we have to distinguish two cases: An individual state $\varphi^{A_1 A_2 A_3}$ can result in “small” correlation between J_1^k and $A_2 B_3$ – but then Lemma 4 above limits the correlation between J_1 and $A_2 B_3$, making input A_1 of almost no use (Proposition 5 below). On the other hand, if there is “large” correlation, we can use the J_ℓ to break up the input state into an ensemble acting only on $A_2 A_3$ with at least the same contribution to the Holevo quantity (Proposition 6). For the following two Propositions, we do the same thing as above, keeping a record of J_1^k , together with the output state on $B_0 B_2 B_3$. However, notice that after a unitary U is applied to A_1 and then the measurements on the $A_{1\ell}$ performed, $A_2 A_3$ collapses into a state $\varphi_{j_1^k}(U)^{A_2 A_3}$ with probability $p_{j_1^k}(U)$:

$$p_{j_1^k}(U) \varphi_{j_1^k}(U)^{A_2 A_3} = \langle j_1 \dots j_k | U^{A_1} \varphi U^{A_1 \dagger} | j_1 \dots j_k \rangle.$$

With respect to which *input* state entropic quantities are to be interpreted is indicated by adding that input state as a subscript (unless it is φ).

Proposition 5 For the joint channel $\mathcal{T}_N^k \otimes \mathcal{E}$ with input state $\varphi^{A_1 A_2 A_3}$ and $\sigma^{A_1 A_2 B_3} = (\text{id}_{A_1 A_2} \otimes \mathcal{E}) \varphi^{A_1 A_2 A_3}$, if $I(J_1^k; A_2 B_3 | B_0) < 4 \log |E|$, then

$$H(B_1 B_2 B_3 | B_0) > H(B_2 B_3) + \log |E| - \delta(k),$$

where $\delta(k) = \frac{1}{k}(5 + 4 \log |E|)$.

Proof. We start by invoking Lemma 2 to the entropy of $J_\ell \dots J_k$: this yields

$$H(J_\ell \dots J_k | B_0) \geq 2(k - \ell + 1) \log |E| - \frac{2}{|E|^{2(\ell-1)}},$$

which implies $I(J_\ell; J_{\ell+1}^k | B_0) \leq \frac{2}{|E|^{2(\ell-1)}}$.

Now, with Lemma 4 and using the assumption that $I(J_1^k; A_2 B_3 | B_0) < 4 \log |E|$, we find

$$\begin{aligned} I(J_1; A_2 B_3 | B_0) &\leq \frac{1}{k} I(J_1^k; A_2 B_3 | B_0) \\ &\quad + \frac{1}{k} \sum_{\ell=1}^{k-1} I(J_\ell; J_{\ell+1}^k | B_0) \\ &\leq \frac{1}{k} 4 \log |E| + \frac{1}{k} \sum_{\ell=1}^{\infty} \frac{2}{|E|^{2(\ell-1)}} \\ &= \frac{1}{k} 4 \log |E| + \frac{1}{k} \frac{2}{1 - 1/|E|^2} \\ &\leq \frac{1}{k} (3 + 4 \log |E|). \end{aligned} \tag{16}$$

For the given input state $\varphi^{A_1 A_2 A_3}$, let us write the quantum state of the system $J_1 A_2 B_3$ as $\omega^{J_1 A_2 B_3}(U)$, and the output state on $B_1 B_2 B_3$ as $\phi^{B_1 B_2 B_3}(U)$. We also denote the CPTP quantum operation mapping $J_1 A_2 B_3$ to $B_1 B_2 B_3$ in our setting of $\mathcal{T}_N^k \otimes \mathcal{E}$ as \mathcal{R} . Then

$$\begin{aligned} \mathcal{R}(\omega^{J_1 A_2 B_3}(U)) &= \phi^{B_1 B_2 B_3}(U), \\ \mathcal{R}\left(\frac{1}{|E|^2} \mathbb{1}^{J_1} \otimes \omega^{A_2 B_3}\right) &= \frac{1}{|E|} \mathbb{1}^{B_1} \otimes \phi^{B_2 B_3}. \end{aligned}$$

By straightforward calculation, and using Eq. (16) and Lemma 2 once more, we have

$$\begin{aligned} \mathbb{E}_U D\left(\omega^{J_1 A_2 B_3}(U) \middle\| \frac{1}{|E|^2} \mathbb{1}^{J_1} \otimes \omega^{A_2 B_3}\right) \\ &= 2 \log |E| + I(J_1; A_2 B_3 | B_0) - H(J_1 | B_0) \\ &\leq 2 \log |E| + \frac{1}{k} (3 + 4 \log |E|) - \left(2 \log |E| - \frac{2}{|E|^{2(k-1)}}\right) \\ &\leq \frac{1}{k} (5 + 4 \log |E|), \end{aligned}$$

where $D(\rho \| \sigma) := \text{Tr}(\rho(\log \rho - \log \sigma))$ is the quantum relative entropy. By the Lindblad-Uhlmann theorem [28], the quantum relative entropy is monotonic under completely positive quantum operations. As a result, we obtain

$$\begin{aligned} \log |E| - H(B_1 B_2 B_3 | B_0) + H(B_2 B_3) \\ &= \mathbb{E}_U D\left(\phi^{B_1 B_2 B_3}(U) \middle\| \frac{1}{|E|} \mathbb{1}^{B_1} \otimes \phi^{B_2 B_3}\right) \\ &\leq \mathbb{E}_U D\left(\omega^{J_1 A_2 B_3}(U) \middle\| \frac{1}{|E|^2} \mathbb{1}^{J_1} \otimes \omega^{A_2 B_3}\right) \\ &\leq \frac{1}{k} (5 + 4 \log |E|), \end{aligned}$$

which is exactly to say

$$H(B_1 B_2 B_3 | B_0) \geq H(B_2 B_3) + \log |E| - \frac{1}{k} (5 + 4 \log |E|),$$

and we are done. \square

Proposition 6 For the joint channel $\mathcal{T}_N^k \otimes \mathcal{E}$ with input state $\varphi^{A_1 A_2 A_3}$ and $\sigma^{A_1 A_2 B_3} = (\text{id}_{A_1 A_2} \otimes \mathcal{E})\varphi^{A_1 A_2 A_3}$, if $I(J_1^k; A_2 B_3 | B_0) \geq 4 \log |E|$, then there exists a particular unitary U_0 such that

$$H(B_1 B_2 B_3 | B_0) \geq \sum_{j_1 \dots j_k} p_{j_1^k}(U_0) H_{\varphi_{j_1^k}(U_0)}(B_2 B_3) + \log |E|.$$

Proof. By assumption, we have

$$\begin{aligned} 4 \log |E| &\leq I_{\varphi^{A_1 A_2 A_3}}(J_1^k; A_2 B_3 | B_0) \\ &= I_{\varphi^{A_1 A_2 A_3}}(J_1^k; E B_2 B_3 | B_0) \\ &\leq I_{\varphi^{A_1 A_2 A_3}}(J_1^k; B_2 B_3 | B_0) + 2 \log |E|, \end{aligned}$$

since the discarding of the register E cannot reduce the mutual information by more than $2 \log |E|$. Thus,

$$\begin{aligned} 2 \log |E| &\leq I_{\varphi^{A_1 A_2 A_3}}(J_1^k; B_2 B_3 | B_0) \\ &= H_{\varphi^{A_2 A_3}}(B_2 B_3) - H_{\varphi^{A_1 A_2 A_3}}(B_2 B_3 | J_1^k B_0). \end{aligned} \tag{17}$$

Hence, there must be a special unitary U_0 , such that

$$\begin{aligned} H_{\varphi^{A_1 A_2 A_3}}(B_2 B_3 | J_1^k B_0) &\geq H_{\varphi^{A_1 A_2 A_3}}(B_2 B_3 | J_1^k, B_0 = U_0) \\ &= \sum_{j_1^k} p_{j_1^k}(U_0) H_{\varphi_{j_1^k}(U_0)^{A_2 A_3}}(B_2 B_3). \end{aligned} \tag{18}$$

By the subadditivity of von Neumann entropy, we have

$$\begin{aligned} H_{\varphi^{A_1 A_2 A_3}}(B_1 B_2 B_3 | B_0) &\geq H_{\varphi^{A_2 A_3}}(B_2 B_3) - H(B_1 | B_0) \\ &\geq H_{\varphi^{A_2 A_3}}(B_2 B_3) - \log |E|. \end{aligned} \tag{19}$$

At last, putting Eqs. (17-19) together completes the proof. \square

With that we are now ready for the

Proof of Theorem 1. For an ensemble of input states $\{\lambda_x, \varphi_x^{A_1 A_2 A_3}\}$ for $\mathcal{T}_N^k \otimes \mathcal{E}$ with average state $\rho = \sum_x \lambda_x \varphi_x$, we divide it up into two classes according to the above cases:

$$\begin{aligned} \mathcal{G} &:= \{x | I_{\varphi_x}(J_1^k; A_2 B_3 | B_0) \geq 4 \log |E|\}, \\ \mathcal{L} &:= \{x | I_{\varphi_x}(J_1^k; A_2 B_3 | B_0) < 4 \log |E|\}. \end{aligned}$$

Then,

$$\begin{aligned} \chi_{\{\lambda_x, \varphi_x\}}(\mathcal{T}_N^k \otimes \mathcal{E}) &= H_\rho(B_1 B_2 B_3 | B_0) \\ &\quad - \sum_{x \in \mathcal{G}} \lambda_x H_{\varphi_x}(B_1 B_2 B_3 | B_0) \\ &\quad - \sum_{x \in \mathcal{L}} \lambda_x H_{\varphi_x}(B_1 B_2 B_3 | B_0). \end{aligned} \tag{20}$$

By the subadditivity of the von Neumann entropy, the first term is upper bounded

$$\begin{aligned} H_\rho(B_1B_2B_3|B_0) &\leq H_\rho(B_1|B_0) + H_\rho(B_2B_3|B_0) \\ &\leq H_{\rho^{A_2A_3}}(B_2B_3) + \log|E|. \end{aligned} \quad (21)$$

Second, by Proposition 5, we have for $x \in \mathcal{L}$ that

$$H_{\varphi_x}(B_1B_2B_3|B_0) > H_{\varphi_x^{A_2A_3}}(B_2B_3) + \log|E| - \delta(k). \quad (22)$$

Third, by Proposition 6, for $x \in \mathcal{G}$, there is an ensemble decomposition of $\varphi_x^{A_2A_3}$,

$$\varphi_x^{A_2A_3} = \sum_{y=1}^{|E|^{2k}} \mu_{xy} \varphi_{xy}^{A_2A_3},$$

such that

$$H_{\varphi_x}(B_1B_2B_3|B_0) \geq \sum_{y=1}^{|E|^{2k}} \mu_{xy} H_{\varphi_{xy}^{A_2A_3}}(B_2B_3) + \log|E|. \quad (23)$$

Now define the union ensemble of the above states,

$$\mathcal{O} = \{\lambda_x, \varphi_x^{A_2A_3}\}_{x \in \mathcal{L}} \cup \{\lambda_x \mu_{xy}, \varphi_{xy}^{A_2A_3}\}_{x \in \mathcal{G}, y=1, \dots, |E|^{2k}}.$$

Inserting Eqs. (21-23) into Eq. (20) results in

$$\chi_{\{\lambda_x, \varphi_x\}}(\mathcal{T}_{\mathcal{N}}^k \otimes \mathcal{E}) \leq \chi_{\mathcal{O}}(\mathcal{N} \otimes \mathcal{E}) + \delta(k),$$

and we are done. \square

DEPOLARIZING CHANNEL (AUXILIARY MATERIAL)

The difference of $Q_E(\mathcal{D}_q)$ and $C(\mathcal{D}_q)$ of the depolarizing channel \mathcal{D}_q , for several special values of d . The following plot shows q (horizontal axis) against $Q_E(\mathcal{D}_q) - C(\mathcal{D}_q)$ on the vertical axis.

